ANALYSIS

55(286), October, 2019

ISSN 2278–5469 EISSN 2278–5450

DISCOVERY

Evaluating forecasting methods for breast cancer reported cases in Nigeria

Sidiq Okwudili Ben✉, **Dawodu Kikelomo Dolapo**

Department of Geography, College of Education Ikere-Ekiti, Nigeria

✉**Corresponding author**
Department of Geography, College of Education Ikere-Ekiti,
Nigeria
Email: sidiqoben@yahoo.com

Article History
Received: 08 July 2019
Reviewed: 11/July/2019 *to* 29/August/2019
Accepted: 6 September 2019
Prepared: 12 September 2019
Published: October 2019

Citation
Sidiq Okwudili Ben, Dawodu Kikelomo Dolapo. Evaluating forecasting methods for breast cancer reported cases in Nigeria. *Discovery*, 2019, 55(286), 540-556

Publication License
© The Author(s) 2019. Open Access. This article is licensed under a Creative Commons Attribution License 4.0 (CC BY 4.0).

General Note
Article is recommended to print as color digital version in recycled paper.

ABSTRACT

This paper present the evaluation of forecasting techniques for reported breast cancer cases in Nigeria. The most appropriate forecasting method based on accuracy and ease of use (simplicity) to determine the interpolated forecast for the reported cases of breast cancer in Nigeria for the periods of 2008 to 2017 (n = 40) quarterly. Secondary data involved in the reported cases of breast cancer were historical and long on forty quarterly. This study evaluated different forecasting model using reported cases of breast cancer data from the University College Hospital (UCH) Ibadan. Findings revealed that two yearly moving average (n=2) and the yearly moving average (n=3) are most accurate, as they has lowest value of Mean Absolute Deviation (MAD). Also, exponential smoothing, the smoothing constant of 0.9 appears to be the lowest value of Mean Absolute Deviation (MAD). Also, there is

maximum error term and the explanation is very low; hence the prediction or forecast by OLS regression seems to be unreliable and that interpolated forecast produced by OLS regression cannot be suitable for extrapolation with this nature of data.

Keywords: forecasting, evaluation, breast cancer, moving average, exponential smoothing

1. INTRODUCTION

Breast cancer is a significant cause of cancer-related mortality in women all over the world (Agboola *et al*., 2012). The incidence of breast cancer in women is rising across the globe (GLOBOCAN, 2012). A comparison of the World Health Organization's International Agency for Research on Cancer [IARC], latest two versions (GLOBOCAN, 2008 & 2012) showed that the number of new cases increased from 12.7 million in 2008 to 14.1 million cases in 2012. Approximately 1.4 million women were diagnosed with breast cancer in 2012. There are 6.3 million women alive who have been diagnosed in the previous five years (Ferlay *et al*., 2010). Since the 2008 estimates, breast cancer incidence has increased by more than 20% while mortality has increased by 14%.

GLOBOCAN (2012), predict that the total number of new breast cancer cases in women would increase to an alarming number of 22 million by 2025. The mortality rate is expected to rise to 13 million at that rate. Hudis (2014) stated that possible explanations for this observation are fourfold:

a. Many countries now have resources to diagnose and report breast cancer more accurately;
b. Populations are growing;
c. Women around the world are living longer; and
d. There is a global trend towards weight gain and obesity, with a broad adoption of the Western lifestyle and diet.

Hudis (2014) therefore called for applying the strategies that were successfully used in the West to bringing down the mortality rate of breast cancer to the developing countries so as to save millions of lives. In Nigeria, GLOBOCAN (2012) reported 27,304 cases of breast cancer in women in for the year 2014. This figure amounted to 42.2% of all cases of cancer in women at that time, with a mortality of 13,960 cases. This figure amounts to 34.3% of deaths from cancer in women. The five-year prevalence prediction for Nigeria stands at 87,579 cases. This figure amounts to 53.1% of all cancers in women. The data of breast cancer in Nigeria show a significant disparity in the health outcomes of women living with breast cancer.

Maxmen (2012) stated that many factors influence a woman's chances of survival, including how early the tumor was detected and the molecular profile of the tumor. Women diagnosed now are much more likely to survive than women in decades past (Fregene *et al*. 2005; Maxmen, 2012). This is because women are living longer each decade because of improvements in surgery, screening, chemotherapies, hormone and biologic therapies. Tumors that are discovered while still localized, grant patients the best prognosis. As cancer spreads, it often becomes increasingly difficult to cure (Maxmen, 2012). Incidentally, breast cancer can be tracked along the continuum of care at different stages by the application of effective strategies for prevention, early detection, treatment, and care (Yip *et al*., 2012).

With the estimated increase in the burden of breast cancer, urgent action is needed to understand the determinants of these health outcomes so as to arrest this trend. Such understanding can inform programs that will assist in the implementation of evidence-based strategies for prevention, early detection, diagnosis, treatment, and palliation of breast cancer in women, while making the best use of available resources. This study intends to explore the statistical trend of reported cases of breast cancer in Nigeria. Specifically, this study sets out to explore various quantitative forecasting techniques, evaluating the best techniques with the purpose of identifying the most accurate forecast.

According to Nijat, David, Peter, and Peter (2016), selecting the most suitable forecasting technique in planning is quite challenging, it that requires having a detailed record of past data and having a comprehensive analysis of empirical results. Recent research findings reveal that assessing the most suitable method of forecasting out of every other forecasting method is referred to as the performance evaluation of forecasting models; it depends on the accuracy measures adopted.

Evaluating the performance of the forecasting method is very crucial, in the last three decades various accuracy measures have been adopted by many studies as an evaluation criterion. A number of different forecast accuracy measures for both regression and classification problems have been proposed by earlier researchers together with the comments and recommendations on the use of the relevant measures (Mahmou, 1984; Makridakis, 1993; Hyndman and Koehler, 2006; Sokolova and Lapalme, 2009; Power, 2011). The quantitative forecasting techniques evaluated in this study were single moving average, simple exponential smoothing, and trend analysis. The study is limited to the three forecasting methods because of the ease of use. The forecasting methods analyzed include: single moving average (n = 2, n = 3, n =4, n =5, n = 6, n = 7, n = 8, n = 9, n = 10) and simple exponential smoothing

method (α =0.1, α =0.2, α =0.3, α =0.4, α=0.5, α=0.6, α=0.7, α=0.8, α=0.9). The accuracy of the forecasting method were mean Absolute Deviation (MAD), Mean Square Error (MSE), Root Mean Square Error (RMSE) and coefficient of reliability.

The Cancer Statistics Worldwide (2005) documented that worldwide; more than one million new cases of female breast cancer are diagnosed each year, making it the most commonly occurring disease in women, accounting for over 1/3 of the estimated annual 4.7million cancer diagnosis in females and the second most common tumor after lung cancer in both sexes. It is also the most common female cancer in both developed and developing countries with 55% of it occurring in the developing countries. In addition, the annual worldwide incidences had almost doubled since 1975 and the prevalence and incidences increased with increasing age (Althuis, 2005).

Breast cancer is the most common cause of cancer-related deaths worldwide, and case fatality rates are highest in low-resource countries (Anderson *et al*., 2008; Ferlay *et al*., 2010). Over 411,000 deaths result from breast cancer annually, accounting for greater than 1.6% of female deaths from all causes (Anderson *et al*., 2008). The incidence of breast cancer in Nigeria has risen significantly (Jedy-Agba *et al*., 2012). The age-standardized incidence rate for breast cancer in the period from 1960 to 1969 was 13.7 per 100,000. It rose to 24.7 per 100,000 by 1998-1999; more or less a doubling of incidence over four decades or approximately 25% increase in rate per decade. The rate in 2009-2010 was 54.3 per 100,000. This represents a 100% increase in the last ten years (Jedy-Agba *et al*., 2012).

Despite the threat that breast cancer poses to public health especially in sub-Saharan Africa, few countries in the region have data on breast cancer incidence (Sylla & Wild, 2012). Most of the breast cancer incidence data in Sub-Sahara Africa in recent times were based on reports from registries in The Gambia, Zimbabwe and Uganda (Curado *et al*., 2011). Jedy-Agba *et al*. (2012) reported that the incidence rate of breast cancer in their study was higher than that reported by GLOBOCAN's (2008) estimate of 38.7% per 100,000. According to Forouzanfar (2011), the reported increasing incidence may be real, due to the prevalence of risk factors for these cancers.

Jedy-Agba *et al*. (2012) highlighted the need for high-quality regional cancer registries to serve a vast country like Nigeria to adequately inform policy and allocation of resources for breast cancer treatments. Cancer registries play a significant role in the design and monitoring of disease control activities and policies. Population-based cancer registries are the primary source of information in developed countries like Canada. In developed countries, the health care infrastructure enables the registration of quality cancer data. In low and middle-income countries, where medical facilities are limited or scarce, cancer registration data may be of low quality. According to Curado *et al*. (2009), high-quality data are necessary to guide cancer care and improvement of identified goals.

Afolayan *et al*. (2012) also alluded to the poverty of data and sparse literature review on the trends of breast cancer in Nigeria due to few existing cancer registries most of which are either hospital-based or pathology-based instead of the preferred population-based cancer registries. According to Boyle and Levin (2008), looking ahead with the rapidly rising cancer burden in low and middle-income countries, more high-quality incidence data are needed from regions and countries to establish the breast cancer burden and to monitor its evolution particularly in response to cancer control and care activities (Boyle & Levin, 2008). Many studies (Afolayan *et al*., 2008; 2012; GLOBOCAN, 2012; Jedy-Agba *et al*., 2012) consistently reported and predicted increases in breast cancer incidence and mortality for Nigeria. In developed country like Canada, the Canadian Cancer Society reported that fewer Canadian women are dying from breast cancer than in the past. The Canadian Cancer Society reported a decrease by 42 percent since the peak in 1986 (Canadian Cancer Statistics, 2014). According to the Canadian Cancer Society, women in Canada, who are diagnosed with breast cancer, are living longer than ever before, with 5-year survival rates of 88 percent (Canadian Cancer Statistics, 2014). The situation in Nigeria is not the same.

There is a prediction of more than a 100% increase in incidence and mortality rates of breast cancer in Nigeria by 2030 (Jedy-Agba et al., 2012; Sylla et al., 2012). With the disparities in outcomes in breast cancer between developed and developing countries, it is imperative that action be taken to understand the causes of these differences and address them appropriately. Despite the reported cases of poor prognosis of breast cancer, there is a lack of research evidence precisely detailing the determinants of the observed high mortality rate, particularly in Nigeria. Afolayan et al. (2012) likewise noted the same observation. The current research study intends to make available quality data that can be used as baseline information to guide breast cancer care. This study will predict the occurrence of breast cancer in Nigeria from year 2019 to year 2040. It is expected that this prediction will prepare the concerned stakeholders into improving breast health care for women most especially those living with breast cancer in Nigeria.

The aim of this study is to analyze the forecast of reported cases of breast cancer in Nigeria within a period of ten years. The specific objectives of this research are to determine the interpolated forecast for reported cases of breast cancer in Nigeria using single moving average with different moving averages; to establish the interpolated forecast for reported cases of breast cancer in Nigeria using simple exponential smoothing with different smoothing constants; to examine the interpolated forecast for reported

cases of breast cancer in Nigeria using ordinary least-squares (OLS) linear regression analysis; and to evaluate the interpolated forecast that will be most suitable for extrapolation. Due to the financial constraints, time factor and scarcity of data, this research was unable to gather secondary data from health centers in the whole Nigeria. Therefore, the research was limited University College Hospital (UCH) Ibadan.

2. METHODOLOGY

This study evaluated different forecasting model using reported cases of breast cancer data from the University College Hospital (UCH) Ibadan. Yearly data from 2008 to 2017 were collected and used to forecast. The forecast model used in the analysis included single moving average method ($n = 2$, $n = 3$, $n = 4$, $n = 5$, $n = 6$, $n = 7$, $n = 8$, $n = 9$, $n = 10$), simple exponential smoothing method ($\alpha = 0.1$, $\alpha = 0.2$, $\alpha = 0.3$, $\alpha = 0.4$, $\alpha = 0.5$, $\alpha = 0.6$, $\alpha = 0.7$, $\alpha = 0.8$, $\alpha = 0.9$) and Ordinary Least Square (OLS) regression The most appropriate forecasting method was determined on the basis of accuracy. In this research, the common accuracy method was used: mean absolute deviation (MAD).

Single Moving Average Method

The single moving average method involves calculating the average of observations and then employing that average as the predictor for the next period (Adeniran and Stephens, 2018). The moving average method is highly dependent on n, the number of terms selected for constructing the average.

The equation is as follows: $F_{t+1} = (Y_t + Y_{t-1} + Y_{t-2} + \ldots + Y_{t-n+1})/n$ ....*Equation 1*

Where:

F_{t+1} = the forecast value for the next period;

Y_t = the actual value at period t;

n = the number of terms in the moving average

The optimal n value can be determined by the interactive model in the smallest error. In some method, the general approach has been to use MSE. In this study, the value of n would be 2, 3, 4, 5, 6, 7, 8, 9, and 10.

Exponential Smoothing Method

Exponential smoothing is the frequently encountered forecasting technique which largely overcomes the limitations of moving averages. This method involves the automatic weighting of past data with weights that decrease exponentially with time, i.e. the most current values receive the greatest weighting and the older observations receive a decreasing weighting. The exponential smoothing technique is a weighted moving average system and the underlying principle is that the New Forecast = Old forecast + a proportion of the forecast error (Lucey, 2007). The smoothing constant (α) can be between 0 and 1. The higher value of α (i.e. the nearer to 1), the more sensitive the forecast becomes to current conditions; whereas the lower the value, the more stable the forecast will be by reacting less sensitively to current conditions (Lucey, 2007; Adeniran and Stephens, 2018; Adeniran, Kanyio, and Owoeye, 2018). Lucey (2007) state that an approximate equivalent of exponential smoothing constants (alpha values: α) and a number of period moving average respectively. The equation for the simple exponential smoothing method is:

$F_{t+1} = \alpha Y_{t-1} + (1 - \alpha) F_{t-1}$ .... *Equation 2*

Where:

F_{t+1} = forecast value for the next period,

F_{t-1} = last period forecast,

Y_{t-1} is the last period actual value,

α= the smoothing constant ($0 < \alpha < 1$)

The accuracy of the simple exponential smoothing method strongly depended on the optimal value of (α). In this study, the value of α will be 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, and 0.9.

OLS Regression Method

The most widely used mathematical method for performing both time-series and the causal quantitative forecast is regression analysis. This applies specific mathematical formulas to estimate forecast equations. These equations may then be used to forecast future activity by applying the equations to independent variables that may occur in the future. Regression equations come in many forms. The most common regression equation is one that represents a straight line. The method used to estimate the equation of a straight line that best represents either historical trends or causal relationships is known as ordinary least-squares (OLS) linear regression analysis.

Although based in sophisticated theories of statistics and calculus, OLS linear regression analysis tools are readily available on most personal computer spreadsheet software, SPSS, SAS, and a variety programming languages that may be used to create custom regression models. All that is required of the forecaster is to collect appropriate data, enter the data into a software program, and apply the regression tool. Although applying data to today's regression tools is quite simple, proper interpretation and use of regression results require at least a fundamental knowledge of regression modeling from a theoretical perspective. It is suggested that anyone who will actively participate in performing or interpreting quantitative forecast results, such as those found from regression analysis, seek additional knowledge in statistical modeling. This study adopts time-series forecast using OLS linear regression for further forecasting tools.

For model specification, the regression line is stated as Y = a + Bx................ (*Equation 3*) which is used to generates interpolations and extrapolations.

Y = Breast cancer reported (Dependent variable),

a = Intercept.

B = Parameter/ slope

x = Time (Independent variable)

The difference between raw data (Y) and the interpolations can be seen on the line graph, and calculated by the coefficient of determination. If the difference is wide, it will result in a low coefficient of determination. The implication is that there will be no need for extrapolation; as the extrapolated forecasts cannot be reliable. But if the difference is minimal, it will result in a high coefficient of determination. The implication is that the interpolated forecast will lead to extrapolations.

For the purpose of achieving the regression line, the following equations must be achieved:

$\Sigma y = na + \Sigma xb$................ (*Equation 4*)

$\Sigma xy = \Sigma xa + \Sigma x^2 b$................ (*Equation 5*)

Specification for Forecasting Error

a. Mean Absolute Deviation

Mean Absolute Deviation (MAD) is a common method for measuring overall forecast error. The value is computed by dividing the sum of the absolute values of the individual forecast error by the sample size (the number of forecast periods). The equation is:

$$\mathrm{MAD} = \frac{\sum_{t=1}^{n}(\mathrm{Yt} - \mathrm{Ft})}{\mathrm{n}}$$

Where:

Y_t = the actual value in time period t

F_t = the forecast value in time period t

n = the number of periods

F_t = the forecast value in time period t

n = the number of periods

b. *Coefficient of determination (R^2)*

The difference between the true line and the observed line can be seen on the line graph and calculated by the coefficient of determination R^2. If the difference of lines otherwise referred to as error term is wide, it will result to a low coefficient of determination. The implication is that the predictions of the forecast cannot be reliable. But if the error term is minimal, it will result to a high coefficient of determination. The implication is that the predictions of forecast will be reliable.

$$R^2 = \frac{(YE - \bar{Y})^2}{(y - \bar{Y})^2}$$

Where:

$\bar{Y} = \frac{\Sigma y}{n}$; YE = Forecast (Y)

Evaluation of Forecasting Method

In this study, the most appropriate forecasting method was selected on the basis of the level of accuracy and ease of use. The various forecasting method used to forecast, the accuracy of the forecasting method was assessed using mean absolute deviation (MAD), and Coefficient of determination is the reliability of forecast derived from interpolated forecast. The difference between the extrapolated forecast and interpolated forecast can be seen on the line graph often referred to as coefficient of determination R^2. If the difference is wide, it will result to a low coefficient of determination which implies that the predictions of extrapolated demand

forecast cannot be reliable. But if minimal, it will result to a high coefficient of determination which implies that the predictions of extrapolated forecast will be reliable.

3. RESULTS AND DISCUSSIONS

The results will be in the form of the objectives as stated: to determine the interpolated forecast for reported cases of breast cancer in Nigeria using single moving average with different moving averages; to establish the interpolated forecast for reported cases of breast cancer in Nigeria using simple exponential smoothing with different smoothing constants; to examine the interpolated forecast for reported cases of breast cancer in Nigeria using ordinary least-squares (OLS) linear regression analysis; and to evaluate the interpolated forecast that will be most suitable for extrapolation.

From Table 1 and Figure 1, there is irregular fluctuation of trend regarding breast cancer occurrence in Nigeria. Also there seems to be high rate of breast cancer occurrence in the second and third quarters of year 2008; second quarter of year 2009; second and third quarters of year 2010; second and third quarters of year 2011; third quarter of year 2012; second and third quarters of year 2013, first and second quarters of year 2014; first quarter in year 2015; first quarter in the year 2016; and second, third quarters in year 2017. It is important to note that because of the unpredictability of trend displayed on line graph, there is need for critical analysis to determine a suitable method of forecast.

Table 1 Record of breast cancer from year 2008 to 2017 in Nigeria

YEAR	Quarter	No of Cancer Victims	Percentage Change
YR 2008	Q1	108	
	Q2	128	15.625
	Q3	157	18.47134
	Q4	108	-45.3704
YR2009	Q1	99	-9.09091
	Q2	105	5.714286
	Q3	74	-41.8919
	Q4	96	22.91667
YR 2010	Q1	145	33.7931
	Q2	128	-13.2813
	Q3	152	15.78947
	Q4	108	-40.7407
YR 2011	Q1	90	-20
	Q2	118	23.72881
	Q3	116	-1.72414
	Q4	83	-39.759
YR 2012	Q1	93	10.75269
	Q2	92	-1.08696
	Q3	120	23.33333
	Q4	97	-23.7113
YR 2013	Q1	112	13.39286
	Q2	99	-13.1313
	Q3	100	1
	Q4	78	-28.2051
YR 2014	Q1	130	40
	Q2	121	-7.43802
	Q3	87	-39.0805
	Q4	48	-81.25
YR 2015	Q1	160	70
	Q2	69	-131.884
	Q3	86	19.76744
	Q4	107	19.62617

YR 2016	Q1	114	6.140351
	Q2	95	-20
	Q3	87	-9.1954
	Q4	109	20.18349
YR 2017	Q1	107	-1.86916
	Q2	141	24.11348
	Q3	157	10.19108
	Q4	164	4.268293
TOTAL	N = 40	4388	Percentage Change

Source: Authors computation (2019)

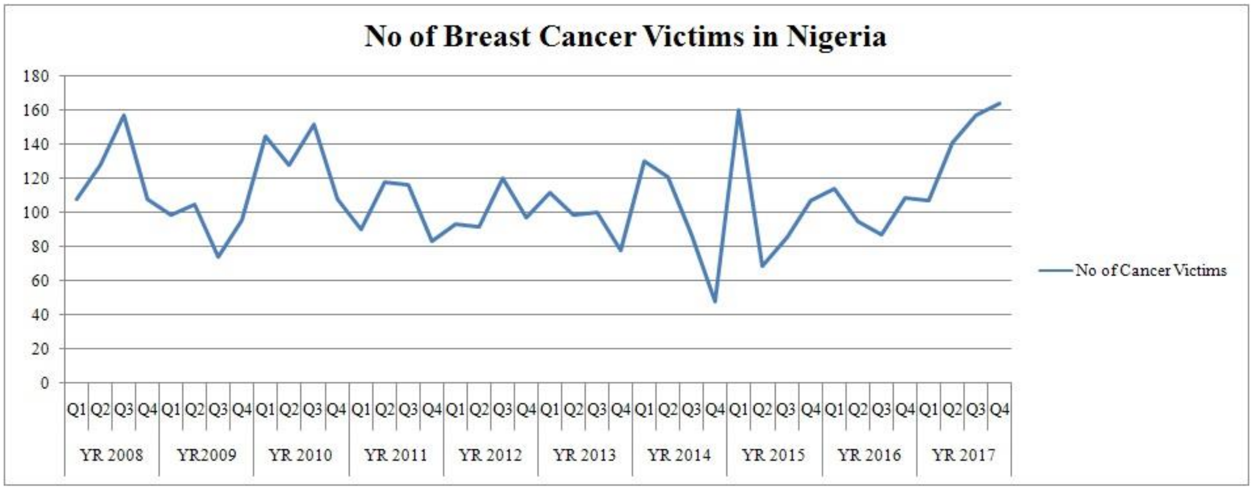

Figure 1 Trend of breast cancer victims in Nigeria

Interpolated Forecast of Breast Cancer Using Single Moving Average

Equation (1) in the model specification has to do with forecasting method using two, three, four, five, six, seven, eight, nine and ten years single moving averages. In order to derive the forecasts, the variables in the equation will be substituted with values, as shown in Table 2 and Figure 2. It was revealed that the lines of interpolated forecast and reported breast cancer have similar trend from the first quarter of 2008 to the fourth quarter of 2017 which might be easily predictable without any critical analysis, but there seems to be a situation of rising and falling which might not be easily predictable without critical analysis. To determine the forecast that seems close to the reported cases is quite difficult without evaluation technique. In the interpolated forecast, there were approximations because forecast cannot be usable in fraction.

Table 2 Interpolated forecast using single moving averages

YEAR	Quarter	No of Cancer Victims	n = 2	n = 3	n = 4	n = 5	n = 6	n = 7	n = 8	n = 9	n = 10
YR 2008	Q1	108									
	Q2	128									
	Q3	157	118								
	Q4	108	143	131							
YR2009	Q1	99	133	131	125						
	Q2	105	104	121	123	120					
	Q3	74	102	104	117	119	118				
	Q4	96	90	93	97	109	112	111			
YR 2010	Q1	145	85	92	94	96	107	110	109		
	Q2	128	121	105	105	104	105	112	114	113	
	Q3	152	137	123	111	110	108	108	114	116	115

	Q4	108	140	142	130	119	117	114	113	118	119
YR 2011	Q1	90	130	129	133	126	117	115	113	113	117
	Q2	118	99	117	120	125	120	113	112	111	111
	Q3	116	104	105	117	119	124	120	114	113	112
	Q4	83	117	108	108	117	119	122	119	114	113
YR 2012	Q1	93	100	106	102	103	111	114	118	115	111
	Q2	92	88	97	103	100	101	109	111	115	113
	Q3	120	93	89	96	100	99	100	107	109	113
	Q4	97	106	102	97	101	104	102	103	108	110
YR 2013	Q1	112	109	103	101	97	100	103	101	102	107
	Q2	99	105	110	105	103	100	102	104	102	103
	Q3	100	106	103	107	104	102	99	102	103	102
	Q4	78	100	104	102	106	103	102	100	101	103
YR 2014	Q1	130	89	92	97	97	101	100	99	97	99
	Q2	121	104	103	102	104	103	105	104	102	100
	Q3	87	126	110	107	106	107	105	107	105	104
	Q4	48	104	113	104	103	103	104	103	105	104
YR 2015	Q1	160	68	85	97	93	94	95	97	97	99
	Q2	69	104	98	104	109	104	103	103	104	103
	Q3	86	115	92	91	97	103	99	99	99	100
	Q4	107	78	105	91	90	95	100	97	98	98
YR 2016	Q1	114	97	87	106	94	93	97	101	98	99
	Q2	95	111	102	94	107	97	96	99	102	100
	Q3	87	105	105	101	94	105	97	96	99	102
	Q4	109	91	99	101	98	93	103	96	95	97
YR 2017	Q1	107	98	97	101	102	100	95	103	97	96
	Q2	141	108	101	100	102	103	101	97	104	98
	Q3	157	124	119	111	108	109	109	106	102	108
	Q4	164	149	135	129	120	116	116	115	111	107

Source: Authors' computation

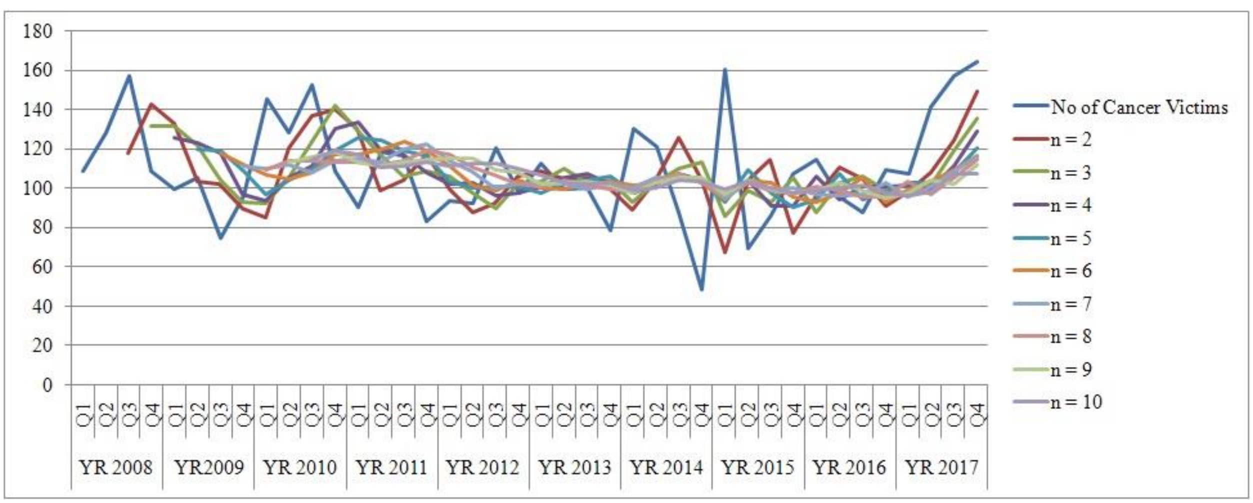

Figure 2 Trend analysis of interpolated forecast of breast cancer using single moving average

Interpolated Forecast of Breast Cancer Using Simple Exponential Smoothing

Equation (2) in the model specification has to do with forecasting method using simple exponential smoothing with smoothing constants of 0.1, 0.2, 0.3, 0.4, 0.5, 0.9, 0.7, 0.8, and 0.9. In order to derive the forecasts, the variables in the equation will be substituted with values, as shown in Table 3 and Figure 3. It was revealed that the lines of interpolated forecast and reported breast

cancer have similar trend from the first quarter of 2008 to the fourth quarter of 2017 which might be easily predictable without any critical analysis, but there seems to be a situation of rising and falling which might not be easily predictable without critical analysis. To determine the forecast that seems close to the reported cases is quite difficult without evaluation technique. In the interpolated forecast, there were approximations because forecast cannot be usable in fraction.

Table 3 Interpolated forecast using simple exponential smoothing

YEAR	Quarter	No of Cancer Victims	α =0.1	α =0.2	α =0.3	α =0.4	α =0.5	α =0.6	α =0.7	α =0.8	α =0.9
YR 2008	Q1	108									
	Q2	128	108	108	108	108	108	108	108	108	108
	Q3	157	110	112	114	116	118	120	122	124	126
	Q4	108	115	121	127	132	138	142	147	150	154
YR2009	Q1	99	114	118	121	123	123	122	120	116	113
	Q2	105	113	115	115	113	111	108	105	102	100
	Q3	74	112	113	112	110	108	106	105	104	105
	Q4	96	108	105	100	96	91	87	83	80	77
YR 2010	Q1	145	107	103	99	96	93	92	92	93	94
	Q2	128	111	111	113	115	119	124	129	135	140
	Q3	152	112	115	117	120	124	126	128	129	129
	Q4	108	116	122	128	133	138	142	145	147	150
YR 2011	Q1	90	115	119	122	123	123	122	119	116	112
	Q2	118	113	114	112	110	106	103	99	95	92
	Q3	116	113	114	114	113	112	112	112	113	115
	Q4	83	114	115	115	114	114	114	115	115	116
YR 2012	Q1	93	111	108	105	102	99	96	93	89	86
	Q2	92	109	105	101	98	96	94	93	92	92
	Q3	120	107	103	99	96	94	93	92	92	92
	Q4	97	108	106	105	105	107	109	112	114	117
YR 2013	Q1	112	107	104	103	102	102	102	101	100	99
	Q2	99	108	106	105	106	107	108	109	110	111
	Q3	100	107	104	104	103	103	103	102	101	100
	Q4	78	106	104	102	102	101	101	101	100	100
YR 2014	Q1	130	103	98	95	92	90	87	85	82	80
	Q2	121	106	105	106	107	110	113	116	120	125
	Q3	87	108	108	110	113	115	118	120	121	121
	Q4	48	105	104	103	103	101	99	97	94	90
YR 2015	Q1	160	100	93	87	81	75	69	63	57	52
	Q2	69	106	106	109	112	117	123	131	139	149
	Q3	86	102	99	97	95	93	91	88	83	77
	Q4	107	100	96	94	91	90	88	86	85	85
YR 2016	Q1	114	101	98	98	98	98	99	101	103	105
	Q2	95	102	101	103	104	106	108	110	112	113
	Q3	87	102	100	100	101	101	100	100	98	97
	Q4	109	100	98	96	95	94	92	91	89	88
YR 2017	Q1	107	101	100	100	101	101	102	104	105	107
	Q2	141	102	101	102	103	104	105	106	107	107

	Q3	157	106	109	114	118	123	127	130	134	138
	Q4	164	111	119	127	134	140	145	149	152	155

Source: Authors' computation

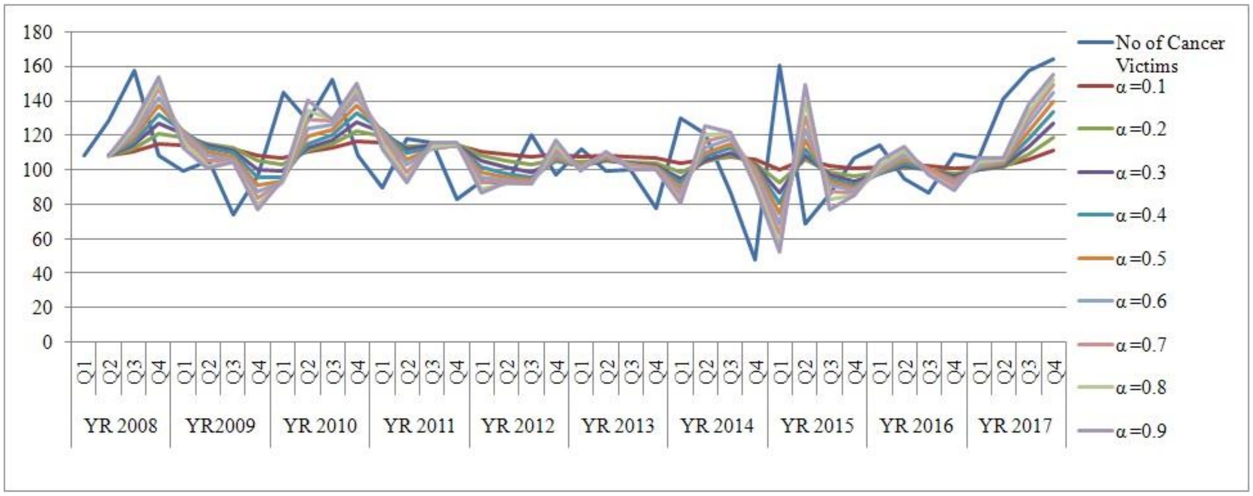

Figure 3 Trend analysis of interpolated forecast of breast cancer using simple exponential smoothing

Interpolated Forecast of Breast Cancer Using Ordinary Least Square (OLS) Regression

Equation (3) in the model specification has to do with forecasting method using OLS regression. In order to derive the interpolated forecasts, the variables will be first substituted into equation 4 and 5. Details are shown in Table 4 and Table 5. From the regression line in Figure 4, it can be seen that the rate of breast cancer occurrence tend rising, hence the need for urgent attention to minimize the occurrence of breast cancer cases.

Table 4 Details of regression arithmetic

YEAR	Number of Quarters (x)	No of Cancer Victims (y)	xy	x^2
YR 2008	1	108	108	1
	2	128	256	4
	3	157	471	9
	4	108	432	16
YR2009	5	99	495	25
	6	105	630	36
	7	74	518	49
	8	96	768	64
YR 2010	9	145	1305	81
	10	128	1280	100
	11	152	1672	121
	12	108	1296	144
YR 2011	13	90	1170	169
	14	118	1652	196
	15	116	1740	225
	16	83	1328	256
YR 2012	17	93	1581	289
	18	92	1656	324
	19	120	2280	361
	20	97	1940	400
YR 2013	21	112	2352	441

	22	99	2178	484
	23	100	2300	529
	24	78	1872	576
YR 2014	25	130	3250	625
	26	121	3146	676
	27	87	2349	729
	28	48	1344	784
YR 2015	29	160	4640	841
	30	69	2070	900
	31	86	2666	961
	32	107	3424	1024
YR 2016	33	114	3762	1089
	34	95	3230	1156
	35	87	3045	1225
	36	109	3924	1296
YR 2017	37	107	3959	1369
	38	141	5358	1444
	39	157	6123	1521
	40	164	6560	1600
TOTAL	820	4388	90130	22140

Source: Author' computation

There are 40 pairs of readings (n = 40)

$\Sigma x = 820$

$\Sigma y = 4388$

$\Sigma xy = 90130$

$\Sigma x^2 = 22140$

All calculations into two decimal places

$\Sigma y = na + \Sigma xb$................ (Equation 4)

$\Sigma xy = \Sigma xa + \Sigma x^2 b$................ (Equation 5)

To obtain the values of a and b, substitute the readings above into equations 4 and 5, and solve simultaneously.

$4388 = 40a + 820b$................ (Equation 4)

$90130 = 4388a + 22140b$................ (Equation 5)

$a = -8.57$

$b = 5.77$

Regression line as shown in the model specification of equation (3) is

$Y = -8.57 + 5.77x$

To forecast, x will be replaced by the number of quarters as shown in table 5 below

Table 5 Determination of interpolated forecast for breast cancer using OLS regression

YEAR	Number of Quarters (x)	No of Cancer Victims (y)	Regression line (Y)
YR 2008	1	108	-2.8
	2	128	2.97
	3	157	8.74
	4	108	14.51
YR2009	5	99	20.28
	6	105	26.05
	7	74	31.82
	8	96	37.59
YR 2010	9	145	43.36

	10	128	49.13
	11	152	54.9
	12	108	60.67
YR 2011	13	90	66.44
	14	118	72.21
	15	116	77.98
	16	83	83.75
YR 2012	17	93	89.52
	18	92	95.29
	19	120	101.1
	20	97	106.8
YR 2013	21	112	112.6
	22	99	118.4
	23	100	124.1
	24	78	129.9
YR 2014	25	130	135.7
	26	121	141.5
	27	87	147.2
	28	48	153
YR 2015	29	160	158.8
	30	69	164.5
	31	86	170.3
	32	107	176.1
YR 2016	33	114	181.8
	34	95	187.6
	35	87	193.4
	36	109	199.2
YR 2017	37	107	204.9
	38	141	210.7
	39	157	216.5
	40	164	222.2

Source: Authors' computation

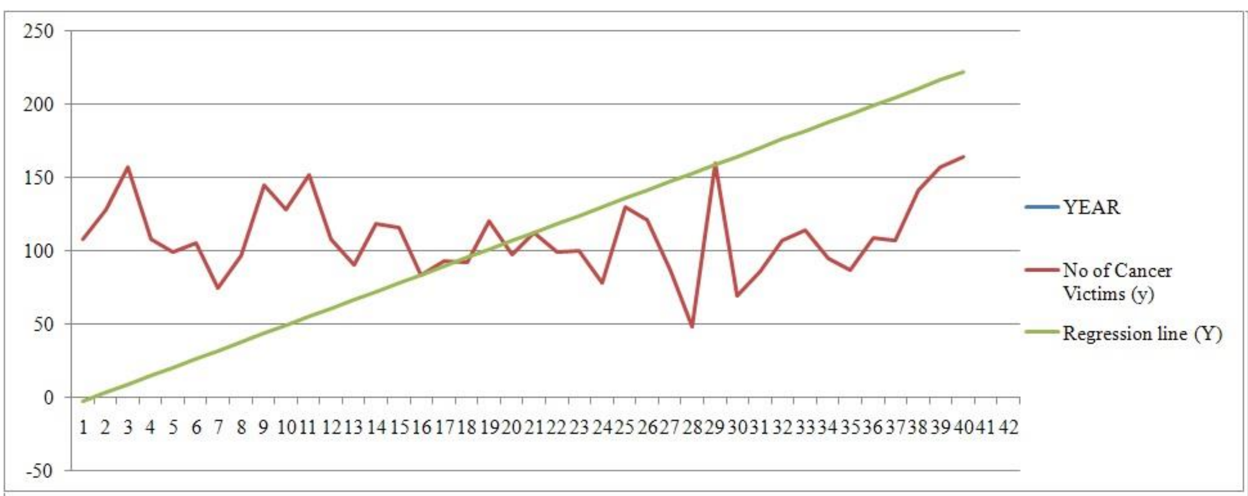

Figure 4 Trend of breast cancer victims and regression line

Evaluation of Interpolated Forecast

In order to get the most suitable interpolated forecast for extrapolation of breast cancer cases reported, there is need to evaluate the most suitable forecast. This is done by comparing yearly single moving averages with exponential smoothing of various smoothing constants, and Ordinary Least Square (OLS) regression using mean absolute deviation (MAD), and Coefficient of determination is the reliability of forecast derived from interpolated forecast.

From Table 6 and Figures 5 and 6 for reported cases of breast cancers in Nigeria, it was revealed that 2 yearly moving average (n=2) and 3 yearly moving average (n=3) are most accurate, as they has lowest value of Mean Absolute Deviation (MAD). Also for exponential smoothing, the smoothing constant of 0.9 appears to be the lowest value of Mean Absolute Deviation (MAD). Hence, it can be deduced from the result that the lower than for the single moving average, the more realistic or reliable the forecast. This corroborates the views of Hsiao (2003), Wooldridge (2001) and Adeniran (2019). Also, the higher the exponential smoothing constant, the more realistic the forecast. This agrees with the study of Hossein (2015); Lucey (2007); Montogomery & (1997); Kahn & Mentzer (1995). It further corroborates the study of Brown (1963) which stated that the higher the values of smoothing constant nearer to 1, the more sensitive the forecast becomes the current condition.

Table 6 Mean Absolute Deviation (MAD) for Single Moving Averages and Simple Exponential Smoothing

Moving Average	n =2	n =3	n =4	n =5	n =6	n =7	n =8	n =9	n =10
MAD	1.3421	1	1.611111	2.457143	2.647059	3.878788	4.28125	3.225806	2.566667
Exponential Smoothing	α = 0.1	α = 0.2	α = 0.3	α = 0.4	α = 0.5	α = 0.6	α = 0.7	α = 0.8	α = 0.9
MAD	2.076923	2.564103	2.512821	2.461538	2.230769	2.051282	1.820513	1.897436	1.615385

Source: Authors' computation

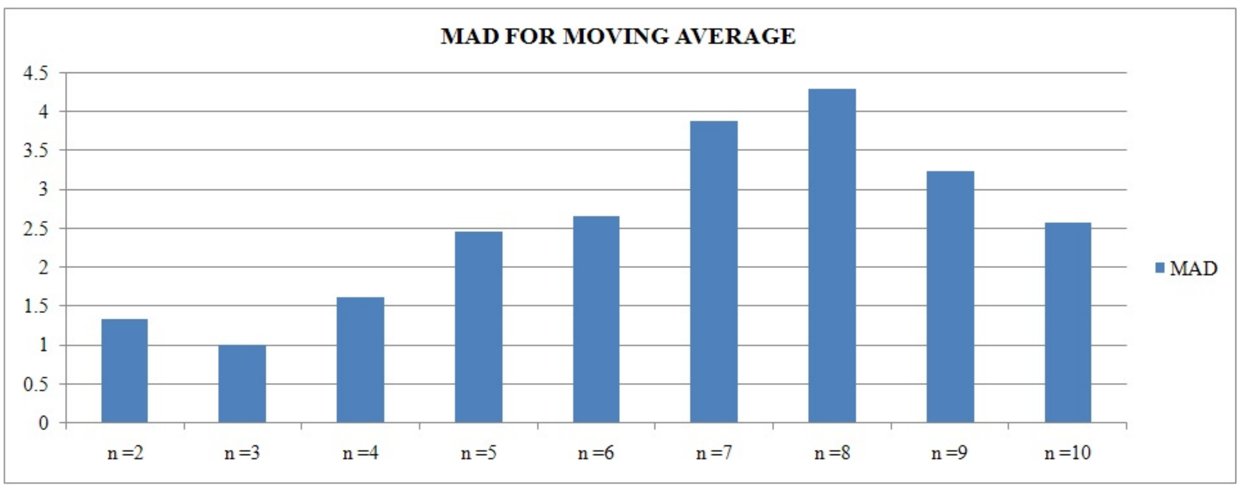

Figure 5 Bar chart showing the Mean Absolute Deviation for single moving average

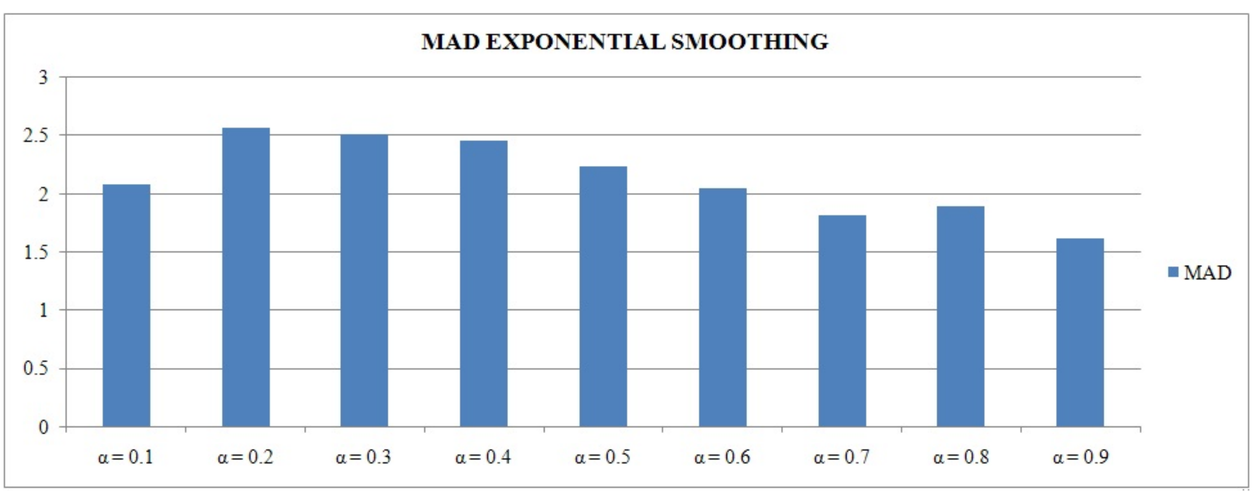

Figure 6 Bar chart showing the Mean Absolute Deviation for simple Exponential Smoothing

Calculation test for Ordinary Least Square (OLS) regression reveals that the coefficient of determination R^2 is 0.154. This implies that the 15.4 percent of variations of the actual breast cancer reported may be predicted by changes in the actual number of periods (quarterly). In other words, the factors other than changes in periods influence the cases of breast cancer reported of (100 – 15.4) percent; 84.6 percent. If the level of explanation be 15.4 percent, then the level of unexplained, or error term, or stochastic disturbance term that is attributed to other factors may be 84.6 percent. This shows that the error term is maximum and very high and the explanation is very low; hence the prediction or forecast by OLS regression seems to be unreliable and that interpolated forecast produced by OLS regression cannot be suitable for extrapolation. This is detailed in Table 7.

Table 7 Calculation of Coefficient of Determination (R^2) for OLS Regression

YEAR	Number of Quarters (x)	No of Cancer Victims (y)	Forecast (YE)	$(YE - \bar{Y})^2$	$(y - \bar{Y})^2$
YR 2008	1	108	-2.8	12656.25	2.89
	2	128	2.97	11391.29	334.89
	3	157	8.74	10192.92	2237.29
	4	108	14.51	9061.136	2.89
YR2009	5	99	20.28	7995.936	114.49
	6	105	26.05	6997.323	22.09
	7	74	31.82	6065.294	1274.49
	8	96	37.59	5199.852	187.69
YR 2010	9	145	43.36	4400.996	1246.09
	10	128	49.13	3668.725	334.89
	11	152	54.9	3003.04	1789.29
	12	108	60.67	2403.941	2.89
YR 2011	13	90	66.44	1871.428	388.09
	14	118	72.21	1405.5	68.89
	15	116	77.98	1006.158	39.69
	16	83	83.75	673.4025	712.89
YR 2012	17	93	89.52	407.2324	278.89
	18	92	95.29	207.6481	313.29
	19	120	101.1	73.96	106.09
	20	97	106.8	8.41	161.29
YR 2013	21	112	112.6	8.41	5.29
	22	99	118.4	75.69	114.49
	23	100	124.1	207.36	94.09
	24	78	129.9	408.04	1004.89
YR 2014	25	130	135.7	676	412.09
	26	121	141.5	1011.24	127.69
	27	87	147.2	1406.25	515.29
	28	48	153	1874.89	3806.89
YR 2015	29	160	158.8	2410.81	2530.09
	30	69	164.5	3003.04	1656.49
	31	86	170.3	3672.36	561.69
	32	107	176.1	4408.96	7.29
YR 2016	33	114	181.8	5198.41	18.49
	34	95	187.6	6068.41	216.09
	35	87	193.4	7005.69	515.29
	36	109	199.2	8010.25	0.49
YR 2017	37	107	204.9	9063.04	7.29

	38	141	210.7	10201	979.69
	39	157	216.5	11406.24	2237.29
	40	164	222.2	12656.25	2948.49

Source: Authors' computation

$\bar{Y} = 109.7$ $(YE - \bar{Y})^2 = 4436.5697$

$(y - \bar{Y})^2 = 684.46$ $R^2 = 0.154$

4. SUMMARY OF FINDINGS

This study identified the most appropriate forecasting method based on accuracy and ease of use (simplicity) to determine the interpolated forecast for the reported cases of breast cancer in Nigeria for the periods of 2008 to 2017 (n = 40) quarterly. Data involved in the reported cases of breast cancer are historical and long on forty quarterly. Due to the financial constraints, time factor and scarcity of data, this research was unable to gather secondary data from health centers in the whole Nigeria. Therefore, the research was limited University College Hospital (UCH) Ibadan. This study evaluated different forecasting model using reported cases of breast cancer data from the University College Hospital (UCH) Ibadan.

The forecast model used in the analysis included single moving average method (n = 2, n = 3, n =4, n =5, n = 6, n = 7, n = 8, n = 9, n = 10), simple exponential smoothing method (α =0.1, α =0.2, α =0.3, α =0.4, α=0.5, α=0.6, α=0.7, α=0.8, α=0.9) and Ordinary Least Square (OLS) regression The most appropriate forecasting method was determined on the basis of accuracy. In this research, the most common accuracy method was used: mean absolute deviation (MAD)

There is irregular fluctuation of trend regarding breast cancer occurrence in Nigeria. Also there seems to be high rate of breast cancer occurrence in the second and third quarters of year 2008; second quarter of year 2009; second and third quarters of year 2010; second and third quarters of year 2011; third quarter of year 2012; second and third quarters of year 2013, first and second quarters of year 2014; first quarter in year 2015; first quarter in the year 2016; and second, third quarters in year 2017. It is important to note that because of the unpredictability of trend displayed on line graph, there is need for critical analysis to determine a suitable method of forecast.

From the study for reported cases of breast cancers in Nigeria, it was revealed that 2 yearly moving average (n=2) and 3 yearly moving average (n=3) are most accurate, as they has lowest value of Mean Absolute Deviation (MAD). Also for exponential smoothing, the smoothing constant of 0.9 appears to be the lowest value of Mean Absolute Deviation (MAD). Hence, it can be deduced from the result that the lower than for the single moving average, the more realistic or reliable the forecast. This corroborates the views of Hsiao (2003), Wooldridge (2001) and Adeniran (2019). Also, the higher the exponential smoothing constant, the more realistic the forecast. This agrees with the the study of Hossein (2015); Lucey (2007); Montogomery & (1997); Kahn & Mentzer (1995). It further corroborates the study of Brown (1963) which stated that the higher the values of smoothing constant nearer to 1, the more sensitive the forecast becomes the current condition.

Calculation test for Ordinary Least Square (OLS) regression reveals that the coefficient of determination R^2 is 0.154. This implies that the 15.4 percent of variations of the actual breast cancer reported may be predicted by changes in the actual number of periods (quarterly). This shows that the error term is maximum and very high and the explanation is very low; hence the prediction or forecast by OLS regression seems to be unreliable and that interpolated forecast produced by OLS regression cannot be suitable for extrapolation.

5. CONCLUSIONS

In line with the objectives, it is concluded that:

i. Two yearly moving average (n=2) and the yearly moving average (n=3) are most accurate, as they has lowest value of Mean Absolute Deviation (MAD).
ii. Exponential smoothing, the smoothing constant of 0.9 appears to be the lowest value of Mean Absolute Deviation (MAD).
iii. There is maximum error term and the explanation is very low; hence the prediction or forecast by OLS regression seems to be unreliable and that interpolated forecast produced by OLS regression cannot be suitable for extrapolation.

Recommendations

From the study, it was recommended that the lower than for the single moving average, the more realistic or reliable the forecast. Also, the higher the values of smoothing constant nearer to 1, the more sensitive the forecast become the current condition.

REFERENCE

1. Adeniran, A. O. (2019). Analytical Study of Evaluating Forecasting Methods in Nigerian Airport. *International Journal of Tourism & Hotel Business Management (IJTHBM)*, 1(1), 32-56.
2. Adeniran, A. O. and Kanyio, O. A. (2018). Long-Term Forecasting of International Air Travel Demand in Nigeria (2018-2028). *American International Journal of Multidisciplinary Scientific Research, 1*(2), 25-31.
3. Adeniran, A. O., Kanyio, A. O., and Owoeye, A. S. (2018). Forecasting Methods for Domestic Air Passenger Demand in Nigeria. *Journal of Applied Research in Industrial Engineering.* Vol. 5, No. 2, Pp. 146-155.
4. Adeniran, A. O, and Stephens, M. S. (2018). The Dynamics for Evaluating Forecasting Methods for International Air Passenger Demand in Nigeria. *Journal of Tourism and Hospitality, 7*(366), 1-11.
5. Afolayan, E. A. O. (2008). Cancer in North Western region of Nigeria: An update, analysis of Zaria cancer registry data, Western Nigeria. *Journal of Medical Science*, 1, 37-43.
6. Afolayan. E, O., Ibrahim, O. O., & Ayilaran, G. T. (2012). An analysis of Ilorin cancer registry statistics. *The Tropical Journal of Health Sciences*, 9, 42-47.
7. Agboola, A. J., Musa, A. A., Wanangwa, N., Abdel-Fatah, T., Nolan, C. C., Ayoade, B. A & Ellis I. O (2012). Molecular characteristics and prognostic features of breast cancer in Nigeria compared with UK women. *Breast Cancer Research and Treatment*, 135(2), 555- 569.
8. Anderson, B. O., Yip, C. H., Smith, R. A., Shyyan, R., Sener, S. F., & Eniu, A., et al. (2008). Guideline implementation for breast healthcare in low-income and middle-income countries: Overview of breast the Breast Health Global Initiative Global Summit 2007. Cancer, 113(Supplement), 2221-2243.
9. Anderson. B. O. (2008).The breast health global initiative: Why it matters to all of us. Oncology, 30(24), 1230-1238.
10. Boyle P & Levin. B (2008). *International Agency for Research on Cancer*. World Cancer Report; Lyon, France.
11. Canadian Cancer Society's Advisory Committee on Cancer Statistics (2014). Canadian Statistics 2014, Toronto, ON: Canadian Cancer Society.
12. Curado, M, P., Edwards, B., Storm, H., Ferly, J., Haenue, M., & Boyle, P. (2011). Cancer incidence in five continents. IARC Scientific Publications No 160.
13. Curado, M. P., Volti, L., & Sortino-Rachoo, A. M. (2009). Cancer registration data and quality indicators in low and middle-income countries: Their interpretation and potential use for the improvement of cancer care. Cancer Causes, 20(5), 751-752.
14. Ferlay J., Shin, H. R., Bray, F., Forman, D., Mathers, C., & Parkin, D. M. (2010). Estimates of worldwide burden of cancer in 2008: GLOBOCAN 2008. Int J Cancer, 127(12), 2893- 2917.
15. Forouzanfar, M. H., Foreman. K. J., Delossantos, A. M. , Lozano, R., Lopez, A. D., Murray, C. J., & Naghaui, M (2011). Breast and cervical cancers in 187 countries between 1980-2010: A Systematic Analysis. Lancet, 378(9801), 1461-1486.
16. Fregene, A., & Newman, L. A. (2005). Breast cancer in sub-Sahara Africa. How does it relate to breast cancer in African - American women? Cancer, 103, 1540-1550.
17. GLOBOCAN (2012). Cancer incidence, mortality and prevalence worldwide. International Agency for Research on Cancer (IARC 2012-2014).
18. Hudis, A. C. (2014). On the rise globally, mortality declines in the US. Cancer discovery. Retrieved from doi:10.1158/2159-8290.CD NB 2014
19. Hyndman, R. J., and Koehler, A. B. (2006). Another look at measures of forecast accuracy. *International Journal of Forecasting,* 22(4): 679-688.
20. International Agency for Research on Cancer. GLOBOCAN 2008. Cancer fact sheets. Retrieved from: www.globocan.iarc/fr/fact sheet_cncer.aspX#WOMEN
21. International Agency for Research on Cancer. GLOBOCAN 2012. Cancer fact sheets. Retrieved from: www.globocan.iarc/fr/fact sheet_cncer.aspX#WOMEN
22. Jedy-Agba, E., Curado, M. P., Ogunbiyi, O., Oga, E., Fabowale, T., Igbinoba, F., & Adebamowo, C. A. (2012). Cancer incidence in Nigeria: A report from population-based cancer registries. Cancer Epidemiology, 36(5), e271-e278.
23. Jedy-Agba, E. E., Curado, M. P., Oga, E., Samaila, M. O., Ezeome, E. R., Obiorah, C., & Adebamowo, C. A. (2012). The role of hospital-based cancer registries in low and middle income countries-The Nigerian Case Study. Cancer Epidemiol, 36(5), 430-435.
24. Hossein, A. (2015). Forecasting by smoothing techniques. Retrieved from: https://home.ubalt.edu/ntsbarsh/Business-stat/otherapplets/ForecaSmo.htm
25. Hsiao, C. (2003). Analysis of panel data. *Cambridge University Press*. 2nd Edn.
26. Kahn, K.B. and Mentzer, J.T. (1995). Forecasting in consumer and industrial markets. *Journal of Business Forecasting Methods & Systems, 14*, 21-28.
27. Lucey, T. (2007). Quantitative Techniques. Sixth Edition, 2007. Book Power/ELST.
28. Mahmoud, E. (1984). Accuracy in forecasting: A survey. *Journal of Forecasting,* 3(2): 139-159.
29. Makridakis S. (1993). Accuracy measures: theoretical and practical concerns. *International Journal of Forecasting,* 9(4): 527-529.
30. Maxmen, A. (2012). The hard facts for women worldwide. Nature Document: URLhttp://go.galegroup.com/ps/ido?.id=

GALE%7CA2G274351&V=2.1&U England Journal of Medicine, 358, 213-216.

31. Montgomery, D.C. and Johnson, L.A. (1997). Forecasting and time series analysis. *McGraw-Hill* New York.
32. Nijat, M., David, E., Peter, F., and Peter, L. (2016). Evaluating Forecasting Methods by Considering Different Accuracy Measures. *Procedia Computer Science, 95*, 264-271
33. Powers, D. M. (2011). Evaluation: from precision, recall and F-measure to ROC, informedness, markedness, and correlation.
34. Ryu, K, and Sanchez, A. (2003) The Evaluation of Forecasting Method at an Institutional Food Service Dining Facility. *Journal of Hospitality and Financial Management, 11*: 27-45.
35. Sokolova M, Lapalme G. (2009). A systematic analysis of performance measures for classification tasks. *Information Processing & Management*, 45(4): 427-437.
36. Sylla, B. S., & Wild, C. P. (2012). A million Africans a year dying from cancer by 2030. What can cancer research and control offer the continent? International Journal of Cancer, 130(2), 245-250.
37. Wooldridge, J.M. (2001). Econometric analysis of cross section and panel data. *The MIT Press*.
38. Yip, C. P., & Low, W. Y. (2011). Recognizing symptoms of breast cancer as a reason for delayed presentation in Asian women. The psycho-socio-cultural model for breast cancer symptoms appraisal: Opportunities for interventions. Asian Pacific Journal of Breast Cancer Prevention, 12, 101.